National Institute for Health and Care Excellence

Final

Early and locally advanced breast cancer: diagnosis and management

Methods

NICE guideline NG101

Methods

July 2018

This methods chapter was developed by the National Guideline Alliance, hosted by the Royal College of Obstetricians and Gynaecologists

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3008-1

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to produce the update for this guideline.

The remit for this guideline update is to revise the NICE clinical guideline on the diagnosis and management of early and locally advanced breast cancer.

What this guideline covers

Groups that are covered

The guideline update covers people with early and locally advanced breast cancer, including:

- adults (18 and over) with newly diagnosed invasive adenocarcinoma of the breast of any size (T1–T4), with or without spread to locoregional lymph nodes (N0–N3) and with no distant metastases (M0)
- adults (18 and over) with newly diagnosed ductal carcinoma in situ (DCIS)
- adults (18 and over) with Paget's disease of the breast.

Clinical areas that are covered

The guideline update covers the following clinical issues:

- surgery to the breast
- management of the positive axilla
- adjuvant systemic therapy planning
- endocrine therapy for invasive disease
- adjuvant chemotherapy
- adjuvant biological therapy
- adjuvant bisphosphonates
- breast radiotherapy
- post-mastectomy radiotherapy
- neoadjuvant treatment of early and locally advanced breast cancer
- lifestyle.

Note that guideline recommendations will normally fall within licensed indications. Exceptionally, and only if clearly supported by evidence, use outside a licensed indication may be recommended. This guideline will assume that prescribers will use a drug's summary of product characteristics to inform decisions made with individual patients.

For further details please refer to the scope on the NICE website (https://www.nice.org.uk/guidance/gid-ng10016/documents/final-scope).

What this guideline does not cover

Groups that are not covered

The guideline does not cover the following groups:

- adults (18 and over) with invasive adenocarcinoma of the breast and distant metastases (clinical or pathological M1)
- adults (18 and over) with rare breast tumours (for example, angiosarcoma, lymphoma)
- adults (18 and over) with benign breast tumours (for example, fibroadenoma).
- adults (18 and over) with phylloides tumour
- adults (18 and over) with locally recurrent breast cancer or DCIS
- adults (18 and over) with lobular carcinoma in situ (LCIS)
- adults (18 and over) with no personal history of breast cancer and an increased risk of breast cancer due to family history.

Clinical areas that are not covered

This guideline does not cover the following areas:

- identifying people in primary care with suspected early and locally advanced breast cancer and referring them to secondary care
- bisphosphonates used for the prevention or treatment of osteoporosis
- the management of breast cancer and related risks in people with a family history of breast cancer.

The following areas in the published guideline were not updated:

- referral, diagnosis, preoperative assessment and psychological support, including the provision of information
- breast reconstruction techniques
- complications of local treatment and menopausal symptoms
- follow-up.

Recommendations in areas that were not updated were edited to ensure that they meet the current editorial standard, and reflect the current policy and practice context.

Methods

This chapter sets out in detail the methods used to review the evidence and to generate recommendations in the guideline. This guideline was developed using the methods described in Developing NICE guidelines: the manual.

Declarations of interest were recorded according to the 2014 NICE Conflicts of interest policy.

Developing the review questions and outcomes

The 22 review questions developed for this guideline were based on the key areas identified in the guideline update scope (https://www.nice.org.uk/guidance/gid-ng10016/documents/final-scope). They were drafted by the NGA and refined and validated by the committee. They cover all areas of the scope and were signed-off by NICE (see Table 1).

The review questions were based on the following frameworks:

- intervention reviews: population, intervention, comparator and outcome (PICO)
- prediction model performance review: population, intervention, comparator, outcome, timing and setting (PICOTS; as suggested by Debray 2017).

These frameworks guided the development of the review protocols, the literature searching process, the critical appraisal and synthesis of evidence and facilitated the development of recommendations by the committee.

Full literature searches, critical appraisals and evidence reviews were completed for all review questions.

Table 1: Description of review questions

Chapter or section	Type of review	Review question	Outcomes
A. Surgery to the breast	Intervention	Q1.1. Do tumour-free tissue margins wider than 0 mm reduce local recurrence for people with invasive breast cancer and/or ductal carcinoma in situ (DCIS) treated with breast conserving surgery?	Critical • Re-operation rate • Local recurrence rate • Patient satisfaction Important • Overall survival • Disease-free survival • Treatment-related morbidity • Health-related quality of life (HRQoL) • Cosmetic result
B. Management of the positive axilla	Intervention	Q2.1. Is there a subgroup of people who do not need	Critical

Chapter or section	Type of review	Review question	Outcomes
		axillary treatment when the axilla has been found to contain metastatic disease?	• Locoregional recurrence • Treatment-related morbidity • HRQoL Important • Overall survival • Breast cancer specific survival • Rate of adjuvant therapy
	Intervention	Q2.2. What are the best strategies to prevent lymphoedema following axillary intervention?	Critical • Lymphoedema • HRQoL Important • Intervention-related morbidity • Arm and shoulder function • Psychological morbidity
C. Adjuvant systemic therapy planning	Intervention	Q3.1. Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?	Critical • Disease-free survival • Overall survival Important • Treatment-related morbidity
	Prediction model performance	Q3.2. What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	Critical • Tool discrimination (AUROC) • Tool calibration (mortality ratio or survival ratio) • Disease-free survival Important • Accuracy (sensitivity/ specificity) • Overall survival
D. Endocrine therapy for invasive disease	Intervention	Q4.1. What is the optimal duration of adjuvant endocrine therapy for people with oestrogen-	Critical • Treatment-related morbidity

Chapter or section	Type of review	Review question	Outcomes
		receptor positive breast cancer?	• Disease-free survival • Overall survival Important • Compliance/ adherence • Treatment-related mortality • HRQoL
		Q4.2. What is the effectiveness of ovarian suppression in addition to endocrine therapy in pre-menopausal women with oestrogen-positive breast cancer?	Critical • Disease-free survival • Treatment-related morbidity • HRQoL Important • Local recurrence rate • Overall survival • Compliance/ adherence • Treatment-related mortality
	Intervention	Q10.4. What is the role of chemoprevention in women following initial treatment for ductal carcinoma in situ (DCIS)?	Critical • Disease free survival • Local recurrence • Treatment related morbidity Important • HRQoL • Overall survival • Treatment adherence
E. Adjuvant chemotherapy	Intervention	Q5.1. Which people with early and locally advanced breast cancer would benefit from the addition of taxanes to anthracycline-based adjuvant chemotherapy?	Critical • Overall survival • Disease-free survival • Treatment-related morbidity Important • Adequate dose intensity)

Chapter or section	Type of review	Review question	Outcomes
			• Treatment-related mortality • HRQoL/Patient satisfaction
F. Adjuvant biological therapy	Intervention	Q6.1. Which people with T1N0 human epidermal growth receptor 2 (HER2)-positive breast cancers benefit from adjuvant trastuzumab in combination with chemotherapy?	Critical • Disease-free survival • Treatment-related morbidity • Overall survival Important • Treatment-related mortality • HRQoL
G. Adjuvant bisphosphonates	Intervention	Q7.1. What are the indications for using adjuvant bisphosphonates in people with early and locally advanced breast cancer?	Critical • Overall survival • Disease-free survival • Treatment-related morbidity Important • Bone health • Treatment-related mortality • HRQoL
H. Breast radiotherapy	Intervention	Q8.1. What radiotherapy techniques are effective for excluding the heart from the radiation field without compromising coverage of the whole breast target volume for people with early or locally advanced breast cancer?	Critical • Mean heart dose • Target coverage Important • Local recurrence rate • Treatment-related morbidity • Treatment-related mortality
	Intervention	Q8.2. Is there a subgroup of people with early invasive breast cancer who do not need breast radiotherapy after breast-conserving surgery?	Critical • Local recurrence rate • Treatment-related morbidity • HRQoL Important • Overall survival

Chapter or section	Type of review	Review question	Outcomes
			• Disease-free survival • Treatment-related mortality
		Q8.3. Is there a subgroup of women with early invasive breast cancer for whom partial breast radiotherapy is an equally effective alternative to whole breast radiotherapy after breast-conserving surgery?	Critical • Local recurrence rate • Treatment-related morbidity • HRQoL Important • Overall survival • Disease-free survival • Treatment-related mortality • Unplanned additional radiotherapy
	Intervention	Q8.4. What are the indications for radiotherapy to internal mammary nodes?	Critical • Loco-regional recurrence rate • Disease-free survival • Treatment-related morbidity Important • Overall survival • HRQoL
I. Post-mastectomy radiotherapy	Intervention	Q9.1. What are the indications for post mastectomy radiotherapy for people with early and locally advanced breast cancer?	Critical • Loco-regional recurrence rate • Treatment-related morbidity • Overall survival Important • Disease-free survival • Treatment-related mortality • HRQoL
	Intervention	Q9.2. Should the potential need for radiotherapy	Critical • Patient satisfaction

Chapter or section	Type of review	Review question	Outcomes
		preclude immediate breast reconstruction?	• Delay in adjuvant therapy • Complication rates Important • Local recurrence rate • Cosmetic result • HRQoL
J. Neoadjuvant treatment	Intervention	Q10.1. What is the effectiveness of neoadjuvant chemotherapy?	Critical • Local recurrence • Disease-free survival Important • Pathological complete response • Breast-conservation rate • Overall survival • Response rates
	Intervention	Q10.2. Is there a benefit for neoadjuvant endocrine therapy for people with early and locally advanced breast cancer?	Critical • Disease-free survival • Breast conservation rates • Changes in tumour size Important • Overall survival • Local recurrence following surgery • HRQoL
	Intervention	Q10.3. What are the indications for post mastectomy radiotherapy following neoadjuvant systemic therapy?	Critical • Loco-regional recurrence rate • Disease-free survival • Treatment-related morbidity Important • Overall survival • HRQoL
	Intervention	Q10.5. Do people with triple negative or BRCA	Critical

Chapter or section	Type of review	Review question	Outcomes
		germ line mutation with early and locally advanced breast cancer benefit from the addition of a platinum to anthracycline (± taxanes) based neo-adjuvant chemotherapy?	• Pathological complete response rate • Overall survival • Disease-free survival Important • Overall response rate • Adequate dose intensity • Breast conservation rate • Local recurrence rate • Treatment-related morbidity • Treatment-related mortality) • HRQoL
K. Lifestyle	Intervention	Q11.1. What lifestyle changes improve breast cancer-specific outcomes in people treated for early and locally advanced breast cancer?	Critical • Overall survival • Disease-free survival Important • Intervention related morbidity • HRQoL

AUROC: area under the receiver operating characteristic curve; BRCA: BReast CAncer susceptibility gene; DCIS: ductal carcinoma in situ; HER2: human epidermal growth factor receptor 2; HRQoL: health-related quality of life; PR: progesterone receptor

Searching for evidence

Clinical search literature

Systematic literature searches were undertaken to identify all published clinical evidence relevant to the review questions.

Databases were searched using relevant medical subject headings, free-text terms and study type filters where appropriate. Studies published in languages other than English were not reviewed. All searches were conducted in MEDLINE, Embase and The Cochrane Library, with some additional database searching in AMED, PsycINFO and CINAHL for certain topic areas.

Re-run searches were carried out in late September 2017. Re-run searches were not conducted:

- where the initial search was done in September 2017
- for radiotherapy topics (evidence reports H and I; review question 10.3) as the committee advised that it was unlikely that new evidence would have been published for these topics
- for review questions 1.1, 3.1, 6.1 and 10.4 as the committee agreed there was unlikely to be new evidence and/or they had made strong recommendations which were unlikely to be changed.

Any studies added to the databases after the date of the last search (even those published prior to this date) were not included unless specifically stated in the text.

Search strategies were quality assured by cross-checking reference lists of highly relevant papers, analysing search strategies in other systematic reviews and asking the group members to highlight any additional studies. The questions, the study types applied, the databases searched and the years covered can be found in Appendix F in each evidence review chapter.

Searching for grey literature or unpublished literature was not undertaken. Searches for electronic, ahead-of-print publications were not undertaken except for topic 3.2 where the committee was aware of relevant studies in the process of publication.

During the scoping stage, a search was conducted for guidelines and reports on websites of organisations relevant to the topic. All references suggested by stakeholders at the scoping consultation were considered.

Health economics search literature

A global search of economic evidence was undertaken in December 2016 and re-run in September 2017. The following databases were searched:

- MEDLINE (Ovid)
- EMBASE (Ovid)
- Health Technology Assessment database (HTA)
- NHS Economic Evaluations Database (NHS EED).

Further to the database searches, the committee was contacted with a request for details of relevant published and unpublished studies of which they may have knowledge; reference lists of key identified studies were also reviewed for any potentially relevant studies. Finally, the NICE website was searched for any recently published guidance relating to early and locally advanced breast cancer that had not been already identified via the database searches.

The search strategy for existing economic evaluations combined terms capturing the target condition (breast cancer) and, for searches undertaken in MEDLINE and EMBASE, terms to capture economic evaluations. No restrictions on language or setting were applied to any of the searches, but a standard exclusions filter was applied (letters, animals, etc.). Full details of the search strategies are presented in Supplement 1: Health economics.

Call for evidence

No call for evidence was made.

Reviewing clinical evidence

Systematic review process

The evidence was reviewed following these steps.

- Potentially relevant studies were identified for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Full papers were reviewed against pre-specified inclusion and exclusion criteria in the review protocols (in appendix A of each evidence review chapter).
- Key information was extracted on the study's methods, according to the factors specified in the protocols and results. These were presented in summary tables (in each review chapter) and evidence tables (in appendix F of each evidence review chapter).
- Relevant studies were critically appraised using the appropriate checklist as specified in Developing NICE guidelines: the manual.
- Summaries of evidence were generated by outcome (included in the relevant review chapters) and were presented in committee meetings.
- Randomised and non-randomised studies: meta-analysis was carried out where appropriate and results were reported in GRADE profiles (for intervention reviews).
- Model performance studies: data were presented individually by study.

All drafts of reviews were checked by a senior reviewer.

Type of studies and inclusion/exclusion criteria

Systematic reviews (SRs) with meta-analyses were considered the highest quality evidence to be selected for inclusion.

For intervention reviews, randomised controlled trials (RCTs) were included because they are considered the most robust study design for unbiased estimation of intervention effects. Based on their judgement, if the committee believed RCT data were not appropriate or there was limited evidence from RCTs, they agreed to include cohort studies with a comparative group.

For the prediction model performance review, the committee prioritised observational studies.

Posters, letters, editorials, comment articles, unpublished studies and studies not in the English language were excluded. Narrative reviews were also excluded, but individual references were checked for inclusion. Conference abstracts were not routinely included.

For quality assurance of study identification, a 10% random sample of the literature search results was sifted by a second reviewer if:

- the review protocol included non-randomised studies
- the review protocol study inclusion and exclusion criteria were complicated
- the first reviewer was new to the guideline.

Review questions 2.2, 4.2, 8.1, 9.2, 10.1 and 11.1 were dual sifted in this way. Any disagreements were resolved by discussion between the 2 reviewers.

The inclusion and exclusion of studies was based on the review protocols, which can be found in appendix A of each evidence review chapter. Excluded studies and the reasons for their exclusion are listed in appendix L of each evidence review chapter. In addition, the committee was consulted to resolve any uncertainty about inclusion or exclusion.

Methods of combining evidence

Data synthesis for intervention reviews

Pairwise meta-analysis was conducted whenever it could be robustly performed to combine the results of studies using Review Manager 5 (RevMan 5) software.

For binary outcomes, such as occurrence of adverse events, the Mantel-Haenszel method of statistical analysis was used to calculate risk ratios (relative risks, RRs) with 95% confidence intervals (CIs). Where reported, time-to-event data were presented as hazard ratios (HRs).

For continuous outcomes, measures of central tendency (mean) and variation (standard deviation (SD)) are required for meta-analysis. Data for continuous outcomes (such as health-related quality of life score or length of hospital stay) were analysed using an inverse-variance method for pooling weighted mean differences.

Forest plots were generated to visually present the results (please see appendix K of each evidence review chapter).

Stratified analyses were predefined for some review questions at the protocol stage when the committee identified that strata were different in terms of biological and clinical characteristics and the interventions were expected to have a different effect.

Statistical heterogeneity was assessed by visually examining the forest plots, and by considering the chi-squared test for significance at $p<0.1$ or an I-squared inconsistency statistic. Where considerable heterogeneity was present, predefined subgroup analyses were performed. If the heterogeneity still remained, a random effects (DerSimonian and Laird 2015) model was employed to provide a more conservative estimate of the effect. Please note that a random model effect cannot be used for meta-analysis of time to event outcomes reported as observed minus expected events (O – E) and variance in RevMan 5.

Data synthesis of prediction model performance review

To determine the predictive performance of the various prognostic tools, tool calibration and tool discrimination was calculated for each tool.

Tool calibration indicates the accuracy of the prognostic tool to predict an outcome (for example, survival at a given duration of follow-up). This is obtained by calculating the observed: predicted survival ratio.

Tool discrimination indicates the ability of the prognostic tool to discriminate the people developing an outcome (for example, survival at a given duration of follow-

up). This is obtained by calculating the area under the receiver operating characteristic curve (AUROC).

Appraising the quality of evidence

Intervention reviews

GRADE methodology (the Grading of Recommendations Assessment, Development and Evaluation)

For intervention reviews, the evidence for outcomes from the included RCTs was evaluated and presented using GRADE, which was developed by the international GRADE working group.

The software developed by the GRADE working group (GRADEpro) was used to assess the quality of each outcome, taking into account individual study quality factors and the meta-analysis results. The clinical/economic evidence profile tables include details of the quality assessment and pooled outcome data, where appropriate, an absolute measure of intervention effect and the summary of quality of evidence for that outcome. In this table, the columns for intervention and control indicate summary measures of effect and measures of dispersion (such as mean and SD or median and range) for continuous outcomes and frequency of events (n/N; the sum across studies of the number of patients with events divided by sum of the number of completers) for binary outcomes. Reporting or publication bias was taken into consideration in the quality assessment and reported in the clinical evidence profile tables if it was apparent.

The selection of outcomes for each review question was decided when each review protocol was discussed with the committee, and was informed by committee discussion and by key papers.

The evidence for each outcome in the intervention reviews was examined separately for the quality elements listed and defined in Table 2. Each element was graded using the quality levels listed in Table 3.

The main criteria considered in the rating of these elements are discussed below. Footnotes were used to describe reasons for grading a quality element as having serious or very serious limitations. The ratings for each component were summed to obtain an overall assessment for each outcome (Table 4).

Table 2: Description of quality elements in GRADE for intervention reviews

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. High risk of bias for the majority of the evidence decreases confidence in the estimate of the effect.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of results or findings.
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question, or recommendation made, such that the effect estimate is changed. This is also related to applicability or generalisability of findings.

Quality element	Description
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect. Imprecision results if the confidence interval includes the clinically important threshold.
Publication bias	Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies.

Table 3: Levels of quality elements in GRADE

Levels of quality elements in GRADE	Description
None/ no serious	There are no serious issues with the evidence.
Serious	The issues are serious enough to downgrade the outcome evidence by 1 level.
Very serious	The issues are serious enough to downgrade the outcome evidence by 2 levels.

Table 4: Levels of overall quality of outcome evidence in GRADE

Overall quality of outcome evidence in GRADE	Description
High	Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low	Any estimate of effect is very uncertain.

Assessing risk of bias in intervention reviews

Bias is a systematic error, or a consistent deviation from the truth in the results. When a risk of bias is present the true effect can be either under- or over-estimated.

Risk of bias in RCTs was assessed using the Cochrane Risk of Bias Tool (see appendix H in Developing NICE guidelines: the manual).

The possible sources of bias in RCTs in the Cochrane risk of bias tool fit with the following 5 categories: selection bias, performance bias, attrition bias, detection bias and reporting bias.

It should be noted that a study with a poor methodological design does not automatically imply high risk of bias; the bias is considered individually for each outcome and it is assessed whether this poor design will impact on the estimation of the intervention effect.

More details about the tool can be found here:
http://cobe.paginas.ufsc.br/files/2014/10/Cochrane.RCT_.pdf

For observational studies, the methodological quality was assessed using the Newcastle-Ottawa Scale (Wells 2008; see appendix H in Developing NICE guidelines: the manual.

The risk of bias was derived by assessing the risk of bias across 3 domains: selection, comparability and outcome. Studies were given a rating depending on how they performed on each of the domains.

Assessing inconsistency in intervention reviews

Inconsistency refers to unexplained heterogeneity of results of meta-analysis. When estimates of the treatment effect vary widely across studies (that is, there is heterogeneity or variability in results), this suggests true differences in underlying effects. Inconsistency is, thus, only applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled). For outcomes derived from a single study 'no inconsistency' was used when assessing this domain, as per GRADE methodology (Santesso 2016).

Statistical heterogeneity was assessed by visually examining the forest plots, and by considering the chi-squared test for significance at $p<0.1$ and the I-squared inconsistency statistic (with an I-squared value of 50 to 80% indicating potentially serious inconsistency and I-squared value of over 80% indicating very serious inconsistency). Where considerable heterogeneity was present, predefined subgroup analyses were performed. If the heterogeneity still remained, a random effects (DerSimonian and Laird 2015) model was employed to provide a more conservative estimate of the effect. When no plausible explanation for the heterogeneity could be found, the quality of the evidence was downgraded in GRADE by 1 or 2 levels for the domain of inconsistency, depending on the extent of heterogeneity in the results.

Assessing indirectness in intervention reviews

Directness refers to the extent to which the populations, intervention, comparisons and outcome measures are similar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention.

Assessing imprecision and clinical significance in intervention reviews

Imprecision in guidelines concerns whether the uncertainty (confidence interval, CI) around the effect estimate means that it is not clear whether there is a clinically important difference between interventions or not (that is, whether the evidence would clearly support one recommendation or appear to be consistent with several different types of recommendations). Therefore, imprecision differs from the other aspects of evidence quality because it is not really concerned with whether the point estimate is accurate or correct (has internal or external validity). Instead, it is concerned with the uncertainty about what the point estimate actually is. This uncertainty is reflected in the width of the CI.

The 95% CI is defined as the range of values within which the population value will fall on 95% of repeated samples, were this procedure to be repeated. The larger the trial, the smaller the 95% CI and the more certain the effect estimate.

Imprecision in the evidence reviews is assessed by considering whether the width of the 95% CI of the effect estimate is relevant to decision-making, taking each outcome in isolation. This assessment also involves effect size thresholds for clinical importance (the minimally important difference, MID) for benefit and for harm.

If the effect estimate CI includes both clinically important benefit (or harm) and no effect there is uncertainty over which decision to make (based on this outcome alone). The CI is consistent with 2 possible decisions and so this is considered to be imprecise in the GRADE analysis and the evidence is downgraded by 1 level ('serious imprecision').

An effect CI including clinically important benefit, clinically important harm and no effect is consistent with 3 possible decisions. This is considered to be very imprecise in the GRADE analysis and the evidence is downgraded by 2 levels ('very serious imprecision').

Minimally important differences

The literature was searched for established MIDs for the selected outcomes in the evidence reviews. In addition, the committee was asked whether they were aware of any acceptable MIDs in the clinical community.

If no published or acceptable MIDs were identified, the committee considered whether it was clinically acceptable to use the GRADE default MIDs to assess imprecision. For binary outcomes, GRADE default MIDs are RRs of 0.8 and 1.25 (due to the statistical distribution of this measure this means that this is not a symmetrical interval). For continuous outcomes, GRADE default MIDs are half of the SD of the control group.

- For survival outcomes (for example, overall survival or disease-free survival), any statistically significant change was considered by the committee to be clinically important.
- For quality of life, MID values from the literature were used where available:
  - Functional assessment of cancer therapy – General (FACT-G) total: 3-7 points
  - Functional assessment of cancer therapy – Breast cancer (FACT-B) total: 7-8 points
  - Trial outcome index (TOI) of FACT-B: 5-6 points
  - Breast cancer subscale (BCS) of FACT-B: 2-3 points
  - World Health Organization Quality of Life (WHOQOL)-100: 1 point
- For serious adverse events (for example, secondary cancer), any statistically significant change was considered clinically important.
- For all other outcomes, GRADE default MID values were used as a starting point and decisions on clinical importance were then considered based on the absolute risk difference.

Optimal information size

Evaluating the CI is not sufficient to assess imprecision. When there are a small number of events the CI can be narrow but the results may be fragile. Therefore, it is suggested that in addition to considering whether the CI crosses thresholds for MIDs, the optimal information size (OIS), representing the number of patients generated by

a conventional single-trial sample size calculation, should be considered (Schünemann 2013). In statistical hypothesis testing alpha is probability of rejecting the null hypothesis given that it is true and beta is the probability of failing to reject the null hypothesis given that it is false. For continuous outcomes, using the standard alpha and beta values of 0.05 and 0.20 respectively, a total sample size (across both arms) of approximately 400 would be required to detect an effect size of 0.2; therefore if N < 400 for an outcome, the evidence would be considered imprecise and downgraded by 1 level ('serious imprecision'). For binary outcomes, evidence should be considered imprecise and downgraded by 1 level ('serious imprecision') if the total number of events (across both arms) is less than 300. For outcomes where any statistically significant change was considered by the committee to be clinically important, imprecision was rated based on OIS alone; for all other outcomes, imprecision was determined based on the width of the confidence interval and the OIS.

Prediction model performance review

The quality of the studies included in the prediction model performance review were individually assessed using the Critical Appraisal Skills Programme (CASP) tool for clinical prediction rule.

The CASP tool is divided in 3 sections, addressing the following issues.

- Are the results of the study valid?
- What are the results?
- Will the results help locally?

More details about the CASP tool can be found here: http://docs.wixstatic.com/ugd/dded87_a2f74f6cd2f24bd684bb26efe7ad7196.pdf

Evidence statements

Evidence statements are summary statements that are presented after the GRADE profiles, highlighting the key features of the clinical evidence presented. The wording of the evidence statements reflects the certainty or uncertainty in the estimate of effect. The evidence statements are presented by outcome or theme and encompass the following key features of the evidence:

- the quality of the evidence (GRADE rating)
- the number of studies and the number of participants for a particular outcome
- a brief description of the participants
- the clinical significance of the effect and an indication of its direction (for example, if a treatment is clinically significant (beneficial or harmful) compared with another, or whether there is no clinically significant difference between the tested treatments).

Formal consensus methods

Formal consensus methods were used with the committee in instances where relevant clinical evidence was non-existent or insufficient to inform recommendations due to poor quality or lack of evidence for subgroups of interest (review questions 3.1 and 5.1). The modified nominal group technique (Bernstein 1992) was selected due

to its appropriateness for use within the guideline development process. This method, which is the most commonly used in healthcare (Murphy 1998) is effective in quickly obtaining consensus from a range of participants and is transparent, making it possible to trace how a group came to a decision and formed recommendations.

This method required members of the committee to indicate their agreement with a set of statements. The statements were developed by the NGA drawing on available sources of evidence, such as previous guidelines, key papers and discussions with the committee. Agreement with the statements was rated on a 9-point Likert scale where 1 represented strongly disagree, 5 represented neither agree nor disagree and 9 represented strongly agree. Participants had the option of indicating that they had insufficient knowledge in a given area to provide a rating. The ratings were grouped into three categories: 1 to 3 (disagree), 4 to 6 (neither agree nor disagree), or 7 to 9 (agree).

In round 1 of the consensus process, the committee was presented with an overview of the modified nominal group technique, a summary of the available evidence (if any), a consensus questionnaire containing the statements to be rated, and instructions on how to rate the statements in the questionnaire. Committee members were asked to rate their agreement based on their personal opinion of what would constitute best practice, taking into account their expertise, rather than describing current practice. It was emphasised that ratings should be based on agreement with the overall focus of the statement, rather than specific wording. Committee members were also given an opportunity to provide a written comment explaining the reason for any disagreement and how the statement could be modified.

At the subsequent committee meeting committee members were provided with the overall percentage agreement, distribution of responses to each statement, and additional comments. Statements with greater than or equal to 80% agreement were used to inform drafting of recommendations (taking into account comments from the committee members). Statements where there was 60 to 80% agreement were used to inform recommendations if the comments were easy to address with minor amendments, or were redrafted based on the committee's comments, discussed at the committee meeting, and re-rated following the same procedure as in round 1. Statements with less than 60% agreement in round 1 were generally disregarded unless there were obvious and addressable issues identified from the comments. Following round 2 of rating, statements were either used to inform recommendations or disregarded based on percentage agreement.

Economic evidence

The aim of the health economic input to the guideline was to inform the committee of potential economic issues related to management of early and locally advanced breast cancer and to ensure that recommendations represented a cost-effective use of healthcare resources. Health economic evaluations aim to integrate data on healthcare benefits (ideally in terms of quality-adjusted life-years (QALYs)) with the costs of different care options. In addition, the health economic input aimed to identify areas of high resource impact; recommendations which might have a large impact on Clinical Commissioning Group or Trust finances and so need special attention.

Reviewing economic evidence

The titles and abstracts of papers identified through the searches were independently assessed for inclusion using predefined eligibility criteria summarised in Table 5.

Table 5: Inclusion and exclusion criteria for the systematic reviews of economic evaluations

Inclusion criteria
Intervention or comparators according to the scope
Study population according to the scope
Full economic evaluations (cost-utility, cost-effectiveness, cost-benefit or cost-consequence analyses) that assess both the costs and outcomes associated with the interventions of interest
Exclusion criteria
Abstracts with insufficient methodological details
Cost-of-illness type studies

Once the screening of titles and abstracts was complete, full versions of the selected papers were acquired for assessment. The quality of evidence was assessed using the economic evaluations checklist as specified in Developing NICE guidelines: the manual.

Health economic modelling

As well as reviewing the published economic literature, as described above, new economic analysis was undertaken in selected areas prioritised by the committee in conjunction with the health economist. Topics were prioritised on the basis of the following criteria, in accordance with Developing NICE guidelines: the manual:

- the overall importance of the recommendation, which may be a function of the number of patients affected and the potential impact on costs and health outcomes per patient
- the current extent of uncertainty over cost effectiveness, and the likelihood that economic analysis will reduce this uncertainty
- the feasibility of building an economic model.

The following priority areas for de novo economic analysis were agreed by the committee after formation of the review questions and consideration of the available health economic evidence:

- addition of taxanes to anthracycline based adjuvant chemotherapy in people with early and locally advanced breast cancer
- adjuvant trastuzumab in combination with chemotherapy in people with T1N0 HER2 positive breast cancers
- adjuvant bisphosphonates in people with early and locally advanced breast cancer.

The full methods and results of de novo economic analyses are reported in appendix B of each review question that was modelled. When new economic analysis was not prioritised, the committee made a qualitative judgement regarding cost effectiveness

by considering expected differences in resource and cost use between options, alongside clinical effectiveness evidence identified from the clinical evidence review.

Cost effectiveness criteria

NICE's report Social value judgements: principles for the development of NICE guidance sets out the principles that committees should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost effective if any of the following criteria applied (given that the estimate was considered plausible):

- the intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy, or
- the intervention provided clinically significant benefits at an acceptable additional cost when compared with the next best strategy.

The committee's considerations of cost-effectiveness are discussed explicitly under the 'Consideration of economic benefits and harms' headings of the relevant sections.

Resource impact assessment

The resource impact assessment team provides an estimate of the cost or saving ('resource impact') of implementing a guideline.

The resource impact team works alongside guideline committees to support section 7.2 of Developing NICE guidelines: the manual. This states: "Guideline recommendations should be based on the balance between the estimated costs of the interventions or services and their expected benefits compared with an alternative (that is, their 'cost effectiveness'). In general, the committee will want to be increasingly certain of the cost effectiveness of a recommendation as the cost of implementation increases. Therefore, the committee may require more robust evidence on the effectiveness and cost effectiveness of recommendations that are expected to have a substantial impact on resources; any uncertainties must be offset by a compelling argument in favour of the recommendation. The cost impact or savings potential of a recommendation should not be the sole reason for the committee's decision."

The resource impact team follows guideline development from an early stage to identify recommendations that either individually or cumulatively have a substantial impact on resources. The aim is to ensure that a recommendation does not introduce a cost pressure into the health and social care system unless the committee is convinced of the benefits and cost effectiveness of the recommendation.

Resource impact is defined as substantial if:

- the resource impact of implementing a single guideline recommendation in England is more than £1 million per year or
- the resource impact of implementing the whole guideline in England is more than £5 million per year.

As well as costs and savings, the team gives advice to committees on wide-ranging issues such as workforce, capacity and demand, training, facilities and educational implications of the recommendations. It may also advise where responsibility for implementation rests (by identifying commissioners and providers) and who the costs or savings are for (the commissioner or provider).

The overall aim of the team's involvement is to:

- ensure guidelines are supported by good economic evidence if the resource impact is estimated to be substantial
- support future financial planning by profiling the resource impact over the coming 5 financial years if possible
- provide a clear and concise resource impact report and template of the resource impact of implementing a NICE guideline.

There is more information about how resource impact is calculated and how the resource impact team works in the Assessing resource impact process manual: guidelines.

Developing recommendations

Guideline recommendations

Recommendations were drafted on the basis of the committee's interpretation of the available evidence, taking into account the balance of benefits, harms and costs between different courses of action. When clinical and economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on the members' expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs or implications compared with the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues.

The main considerations specific to each recommendation are outlined under the 'Recommendations and link to evidence' headings within each chapter.

For further details please refer to Developing NICE guidelines: the manual.

Research recommendations

When areas were identified for which good evidence was lacking, the committee considered making recommendations for future research. For further details please refer to Developing NICE guidelines: the manual.

Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are responded to in turn and posted on the NICE website at publication. For further details please refer to Developing NICE guidelines: the manual.

Updating the guideline

Following publication, and in accordance with the NICE guidelines manual, NICE will undertake a review of whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update. For further details please refer to Developing NICE guidelines: the manual.

Funding

The NGA was commissioned by NICE to undertake the work on this guideline.

References

Bernstein 1992

Bernstein, S.J., Laouri, M., Hilborne, L.H., Leape, L.L., Kahan, J.P., Park, R., et al. Coronary angiography: a literature review and ratings of appropriateness and necessity. Santa Monica, USA: RAND; 1992. Available from https://www.rand.org/pubs/joint_reports-health/JRA03.html (accessed 26 November 2017)

Debray 2017

Debray, T., Damen, J., Snell, K., Ensor, J., Hooft, L., Reitsma, J. et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ, 356, i6460, 2017.

DerSimonian and Laird 2015

DerSimonian, R., Laird, N. Meta-analysis in clinical trials revisited. Contemporary clinical trials. 2015 Nov 30; 45:139-45.

Murphy 1998

Murphy, M.K., Black, N.A, Lamping, D., McKee, C., Sanderson, C., Askham, J., et al. Consensus development methods, and their use in clinical guideline development. Health Technology Assessment. 1998; 2:i-88.

NICE 2014

National Institute for Health and Care Excellence (NICE), Developing NICE guidelines: the manual (accessed 8 June 2017)

Santesso 2016

Santesso, N., Carrasco-Labra, A., Langendam, M., Brignardello-Petersen, R., Mustafa, R.A., Heus, P., Lasserson, T., Opiyo, N., Kunnamo, I., Sinclair, D. and Garner, P., Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments, Journal of clinical epidemiology, 74, 28-39, 2016.

Schünemann 2013

Schünemann, H., Brożek, J, Guyatt, G., Oxman, A., (eds.). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach (updated October 2013). GRADE Working Group, 2013. Available from gdt.guidelinedevelopment.org/app/handbook/handbook.html (accessed 26 November 2017)

Wells 2008

Wells, G.A., Shea, B., O'Connel,l D., Peterson, J., Welch, V., Losos, M., Tugwel,l P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses, 2008. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 26 November 2017)